

Global Total Least Squares Modeling of Multivariable Time Series

Berend Roorda and Christiaan Heij, *Member, IEEE*

Abstract—In this paper we present a novel approach for the modeling of multivariable time series. The model class consists of linear systems, i.e., the solution sets of linear difference equations. Restricting the model order, the aim is to determine a model with minimal l_2 -distance from the observed time series. Necessary conditions for optimality are described in terms of state-space representations. These conditions motivate a relatively simple iterative algorithm for the nonlinear problem of identifying optimal models. Attractive aspects of the proposed method are that the model error is measured globally, it can be applied for multi-input, multi-output systems, and no prior distinction between inputs and outputs is required. We give an illustration by means of some numerical simulations.

I. INTRODUCTION

THE basic problem in time series modeling is to find a reasonably simple model which gives a sufficiently accurate description of the data. Procedures which have been developed for this problem differ in the specification of the model class and in the way the complexity and accuracy of models is evaluated.

In this paper we will restrict our attention to models described by linear difference equations with fixed coefficients and finite lags. Within this classical setting several modeling procedures have been developed. For an overview, we refer to the textbooks [1], [4], [6]. Well-known examples are the least squares identification of input-output systems in polynomial form and, more generally, the maximum likelihood identification of ARMAX systems, i.e., input-output models where the disturbances follow a moving average process. These and most other methods require that several structural aspects of the model should be specified *a priori*.

- 1) The number of equations, that is, the number of inputs and outputs of the system, and the decomposition of the system variables into inputs and outputs.
- 2) The orders of each of the equations, that is, the so-called structural indices of the system.
- 3) The stochastic properties of the disturbances, in particular the joint correlation structure between inputs, outputs, and disturbances.
- 4) The choice of a canonical parameterization, to avoid problems of nonidentifiable parameters.

Manuscript received July 22, 1993; revised February 5, 1994 and May 25, 1994. Recommended by Associate Editor, S. P. Meyn. The work was supported in part by Grant SC1*-CT92-0779 on System Identification of the Science Program of the Commission of the European Community.

B. Roorda is with the Tinbergen Institute, Erasmus University Rotterdam, Oostmaaslaan 950-952, 3063 DM Rotterdam, Netherlands.

C. Heij is with the Econometric Institute, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, Netherlands.

IEEE Log Number 9406098.

The identification procedure that we propose in this paper differs in some crucial aspects from the methods just described. Our aim is to decompose a given multivariable time series, denoted by w , into two parts, i.e.,

$$w = \hat{w} + \tilde{w} \quad (1)$$

where \hat{w} represents a regular part and \tilde{w} the corresponding deviation. The aim is to keep the approximation error as small as possible, under the condition that the approximating time series \hat{w} is sufficiently regular. To make this more explicit we next describe our notions of regularity and model error.

A time series is called regular if it satisfies linear, time-invariant difference equations of finite lag. Let q denote the number of system variables, p the number of independent equations, and n the total lag, i.e., the sum of the lags of the individual equations. Time series are more regular when $m := q - p$ and n are smaller, i.e., the more equations they satisfy and the smaller the number of initial conditions. This can also be formulated as follows. Define the complexity of a linear, time-invariant, finite dimensional system by the pair (m, n) , where m is the number of system inputs and n the (minimal) number of state variables. A system is called less complex if it has fewer inputs, i.e., unexplained variables, and if it has less states, i.e., initial degrees of freedom. Then a time series is more regular if it can be generated by a less complex system.

The model error is evaluated as follows. For expository reasons we restrict ourselves in this paper to time series which are specified over the infinite time axis \mathbb{Z} and which are square summable, i.e., we assume that $w \in l_2^q$. The main results in this paper can be extended for modeling time series observed over a finite time interval, but we will not treat this issue here to simplify the presentation. The error in approximating an observed time series w by a regular part \hat{w} is measured by the l_2 -norm of the deviation $\tilde{w} = w - \hat{w}$, denoted by $\|\tilde{w}\| := \{\sum_{t=-\infty}^{\infty} \tilde{w}(t)^T \tilde{w}(t)\}^{1/2}$.

In our approach we will assume that the required regularity of the approximating time series \hat{w} has been specified *a priori*. Stated otherwise, we impose an upper bound on the complexity of the system that can generate the approximation. We denote by $\mathcal{B}^{q,m,n}$ the set of all time series that can be generated by systems with m inputs, $q - m$ outputs, and n states, i.e., all time series that satisfy $q - m$ independent linear, time-invariant difference equations with total lag n . Under this restriction we wish to minimize the approximation error as defined above, i.e.,

$$\min\{\|w - \hat{w}\|; \hat{w} \in \mathcal{B}^{q,m,n}\}. \quad (2)$$

Solving this problem for different values of (m, n) gives an impression of the involved trade-off between the required regularity and the resulting approximation error. The search for an acceptable model complexity is facilitated by the fact that the error decreases for increasing (m, n) , as $\mathcal{B}^{q,m_1,n_1} \subset \mathcal{B}^{q,m_2,n_2}$ if $m_1 \leq m_2$ and $n_1 \leq n_2$. Note that this criterion allows for deviations in all the system variables and that the error is not measured locally, e.g., as a prediction error, but as the global l_2 -distance between the observation and the regular part. Therefore we give the name “global total least squares” to the identification criterion (2).

As compared to classical procedures, our formulation of the time series modeling problem involves less *a priori* specifications. With respect to the four structural aspects discussed before, our approach has the following features:

- 1) The number of inputs and outputs is specified *a priori*, but the system variables are all treated alike so that there is no need for a specification of inputs and outputs.
- 2) The structural indexes need not be specified, but only the total lag, and this can be varied easily.
- 3) The problem formulation involves no stochastic specifications, although these may be incorporated by adjusting the norm on l_2 .
- 4) The criterion is nonparametric, so that any representation may be chosen as it suits.

This paper has the following structure. To give some feeling for the global total least squares problem, we first describe the well-known and relatively simple case of static total least squares in Section II. The basis for our modeling theory is the behavioral approach to systems, and this is briefly discussed in Section III. For further treatment of the behavioral approach in systems theory we refer to [10] and [11]. In Sections IV and V we develop a highly structured type of system representations, isometric state representations, which form the cornerstone of our modeling theory. Section VI concerns the question of how to determine an optimal approximation of an observed time series within a given model. This linear optimization problem is solved by a projection algorithm which was introduced in [12]. In Section VII we treat the problem of determining an optimal system. This is a nonlinear optimization problem over a nonconvex set. We propose three model improvement constructions, based on the results in Section VI. These constructions are used in Section VIII to estimate locally optimal models. In Section IX we describe three simulation experiments that illustrate the use of the global total least squares method, and Section X contains some conclusions.

II. STATIC TOTAL LEAST SQUARES

We first consider the well-known case of total least squares in static models. Although it may be somewhat artificial in this case to consider observations in l_2 , i.e., on an infinite time interval, it gives a better introduction for the dynamic case.

Static total least squares involves the approximation of a given time series by a regular one that satisfies linear nondynamic relations. For a required number of independent equations, the objective is to keep the approximation error

as small as possible, i.e., to minimize $\|w - \hat{w}\|$ where w denotes the observed time series and \hat{w} the approximation. If the required number of independent equations is denoted by p , then the regularity of the approximation implies that $R\hat{w} = 0$ for some matrix R of rank p . This means that the approximation \hat{w} has rank at most $m := q - p$. In terms of (2), this class of regular time series is given by $\mathcal{B}^{q,m,0}$, where $n = 0$ corresponds to the exclusion of dynamic equations. This leads to the following formulation of static total least squares.

Definition 2.1 (Static Total Least Squares): For a given time series $w \in l_2^q$ and minimal required number of independent relations $p = q - m$, find a decomposition $w = \hat{w} + \tilde{w}$ with $\hat{w} \in \mathcal{B}^{q,m,0}$ and with $\|\tilde{w}\|$ minimal.

The solution to this problem is given by the singular value decomposition (SVD). We denote the usual Euclidean norm on \mathbf{R}^q by $|\cdot|$ and the induced norm on l_2^q by $\|\cdot\|$.

Proposition 2.2 (Singular Value Decomposition in l_2): Every $w \in l_2^q$ can be decomposed as $w = \sum_{i=1}^q \lambda_i u_i v_i$, with

- 1) $\lambda_1 \geq \dots \geq \lambda_q \geq 0$, called the singular values of w ;
- 2) $u_i \in \mathbf{R}^{q \times 1}$ with $|u_i| = 1$ and $u_i^T u_j = 0$ for $i \neq j$, called the left singular vectors of w ;
- 3) $v_i \in l_2^1$ with $\|v_i\| = 1$ and $v_i \perp v_j$ for $i \neq j$, called the right singular vectors of w .

The singular values are uniquely determined, and if they are distinct, then the singular vectors are also uniquely determined, up to a sign.

Proof: The existence and properties of the SVD for finite matrices are discussed in, e.g., [3]. Let Π denote the empirical covariance matrix of w , i.e., $\Pi = \sum_{t=-\infty}^{\infty} w(t)w(t)^T$. As Π is symmetric, the left singular vectors are equal to the right ones, so let its SVD be given by $\Pi = \sum_{i=1}^q \mu_i u_i u_i^T$. Let $U := [u_1, \dots, u_q]$, then the empirical covariance matrix of $U^T w$ is given by $\text{diag}(\mu_1, \dots, \mu_q)$, so $U^T w$ consists of q orthogonal components of norm $\sqrt{\mu_i}$. It is easily verified that for $v_i := (1/\sqrt{\mu_i})u_i^T w$ and $\lambda_i := \sqrt{\mu_i}$, $w = \sum_{i=1}^q \lambda_i u_i v_i$ is an SVD of w . \square

The SVD solves the static total least squares problem, as follows.

Proposition 2.3 (Optimal Static Approximation): Let $w = \sum_{i=1}^q \lambda_i u_i v_i$ be the SVD of $w \in l_2^q$. Then $\hat{w} := \sum_{i=1}^m \lambda_i u_i v_i$ is the optimal approximation of w in $\mathcal{B}^{q,m,0}$, satisfying the static equations $u_i^T \hat{w} = 0$ for $i = m+1, \dots, q$.

Proof: This result follows immediately from a corresponding property of the SVD for finite matrices, cf. [3]. \square

Example: One of the essential features of total least squares is that all variables are treated in a similar way. For simplicity we consider simulated data from a model with this type of symmetry, namely an errors-in-variables model. This also gives us the opportunity to relate total least squares to other well-known identification methods and to discuss the role of stochastic assumptions. We consider the model

$$u(t) = z(t) + \eta(t), \quad y(t) = \alpha z(t) + \varepsilon(t). \quad (3)$$

The observed variables consist of $w = (u, y)$, and z is an unobserved latent variable and ε and η are unobserved disturbances. Three of the possible methods for the estimation of the parameter α are regression of y on u , regression of u

on y , and total least squares. These methods can be interpreted as maximum likelihood methods, in the case that ε and η are independent white noise processes with variances σ_ε^2 and σ_η^2 and if in addition respectively $\sigma_\eta^2 = 0$, $\sigma_\varepsilon^2 = 0$, or $\sigma_\varepsilon^2 = \sigma_\eta^2$.

As an example, we consider data generated by the model (3) with $\alpha = 1$ and $\sigma_\varepsilon^2 = \sigma_\eta^2 = 0.5$ and where z is a sample of 20 observations of a white noise process with variance one and independent of η and ε . The SVD of the resulting observation is

$$w = 5.64 \begin{pmatrix} 0.72 \\ 0.69 \end{pmatrix} v_1 + 2.31 \begin{pmatrix} -0.69 \\ 0.72 \end{pmatrix} v_2 \quad (4)$$

where v_1 and v_2 are two orthogonal vectors of unit length. According to Proposition 2.3, the optimal static approximation \hat{w} of w of rank 1 is given by the first term in (4), and this satisfies $[-0.69, 0.72]\hat{w} = 0$. This corresponds to an estimated value $0.69/0.72 = 0.96$, which is close to $\alpha = 1$. The corresponding approximation error is $\|w - \hat{w}\| = 2.31$. Regression of y on u yields an estimate of 0.69, and regression of u on y gives 1.36. These results are depicted in Fig. 1.

This example illustrates the fact that the static total least squares (TLS) scheme is in between both regressions. The method is easily adapted for the case $\sigma_\varepsilon^2 \neq \sigma_\eta^2$, by using a weighted l_2 -norm for $\tilde{w} = (\tilde{u}, \tilde{y})$, defined by $\|\tilde{w}\|_\alpha^2 := \alpha^2 \|\tilde{u}\|^2 + \|\tilde{y}\|^2$ with $\alpha = \sigma_\varepsilon/\sigma_\eta$. The regressions correspond to the extremal cases with infinite weight on one of the components. To choose an appropriate weighted l_2 -norm we need information on the relative errors involved in the measured system variables, c.q., the relative weight one attaches to deviations in the different variables. In this paper we will not further address these problems. \diamond

III. GLOBAL TOTAL LEAST SQUARES

In the rest of this paper we are concerned with total least squares in dynamical systems. For this purpose we describe in this section our systems concept and corresponding notions of complexity and misfit.

As stated in the introduction, we consider systems described by difference equations. From the formulation of the global total least squares problem (2), it is obvious that sets of equations with the same solution set are equivalent as they yield the same approximation error for every observation. Hence not the equations themselves, but their solution set is the essential object in our modeling procedure. This is a strong motivation for adopting the behavioral approach to systems, as introduced in [10] and [11], in which a system is defined by the set of time series that are compatible with the system laws. This set is called the behavior of a system. The system laws themselves are considered as a description or representation of the behavior. The properties we impose on difference equations are reflected by the set-theoretic properties of the corresponding behavior. Linear, time-invariant difference equations correspond to linear, shift-invariant behaviors. If in addition the equations have finite lag, it can be decided if a time series belongs to the behavior by scanning it through a finite window. This property is called completeness, which is further explained in Appendix A, (see Definition A.1). We will further restrict the attention

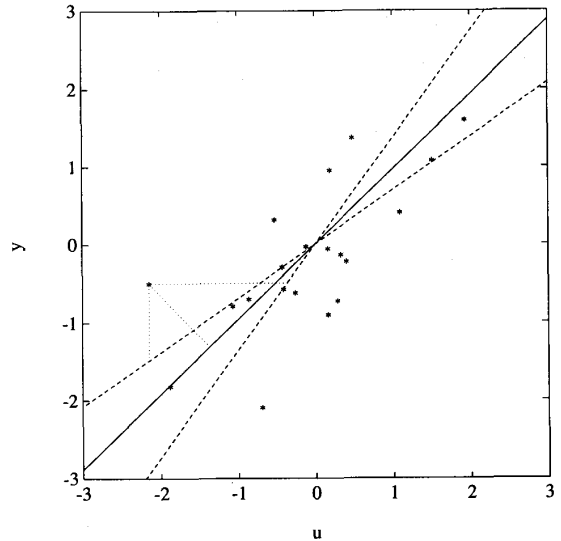


Fig. 1. Scatter diagram of the data w : the solid line denotes the TLS model and the dashed lines the two regression models; the dotted lines indicate which distance is minimized in the three different models.

to behaviors in l_2 . The resulting class of systems is defined as follows.

Definition 3.1 (l_2 -Systems): l_2 -systems are linear, shift-invariant, complete subspaces of l_2^q .

As a measure of the restrictiveness of a set of difference equations we take into account the number of independent equations and their total lag, as described in the introduction. We call the equations independent if the number of equations cannot be reduced without changing the solution set, i.e., if there exists no equivalent smaller set of equations describing the same system. On the set theoretic level of l_2 -systems, we define the complexity of a system in terms of its rank and its degree. The rank of a system is defined as the number of the degrees of freedom at each time instant, which is equal to the number of inputs, and the degree of a system corresponds to the dimension of the state space (see Definition A.3).

Definition 3.2 (Complexity): The complexity of an l_2 -system \mathcal{B} is defined as $c(\mathcal{B}) = (m, n)$, where m denotes the rank of \mathcal{B} and n its degree.

Let $\mathcal{B}^{q,m,n}$ denote the class of l_2 -systems with rank at most m and degree at most n . Then $\mathcal{B}^{q,m,n}$ consists of the solution sets of at least $q-m$ independent linear time-invariant difference equations with the sum of their lags at most n , i.e., $w \in \mathcal{B}^{q,m,n}$ if and only if there exists a system $\mathcal{B} \in \mathcal{B}^{q,m,n}$ with $w \in \mathcal{B}$.

We summarize this in Table I. For further clarification we also relate these concepts to the classical characterization of systems in terms of input-output mappings.

We define the following concept of the misfit of a system.

Definition 3.3 (Misfit): The misfit of an l_2 -system \mathcal{B} with respect to an l_2 -time series w is defined as $d(w, \mathcal{B}) := \inf_{\tilde{w} \in \mathcal{B}} \|w - \tilde{w}\|$.

This leads to the following reformulation of the global total least squares problem (GTLS) as described in the introduction (see (2)).

TABLE I
COMPARISON OF SYSTEM CONCEPTS

Difference Equations	l_2 -behaviours	I/O-mapping
linear	linear subspace	linear
time-invariant	shift-invariant	time-invariant
finite order	l_2 -complete	rational transfer function
$q-m$ indep. equations	rank m	m inputs, $q-m$ outputs
minimal sum of lags is n	degree n	McMillan degree n

Definition 3.4 (GTLS): For an observation $w \in l_2^q$ and given tolerated complexity (m, n) , determine an l_2^q -system $B^* \in \mathcal{B}^{q,m,n}$ such that $d(w, B^*) = \min_{B \in \mathcal{B}^{q,m,n}} d(w, B)$.

This involves a double minimization. The inner minimization, evaluating the misfit $d(w, B)$, amounts to optimization over a linear space. Secondly, we have to determine a system for which the misfit is minimal. This is a nonlinear optimization problem over a nonconvex set.

Leading Example: We describe a simple example, which will be used in the following sections to clarify the introduced general framework. We consider a time series in $\mathcal{B}^{2,1,1}$, that is corrupted by white noise. The regular part w_r consists of two components, u_r and y_r , where u_r is the realization of a white noise process with unit variance, and y_r satisfies

$$y_r(t) = 2/3y_r(t-1) + 2u_r(t) - 2u_r(t-1). \quad (5)$$

The observation w consists of two components u and y , with

$$u(t) = u_r(t) + \eta(t); \quad y(t) = y_r(t) + \varepsilon(t) \quad (6)$$

where η and ε are independent white noise processes with variance 0.25, both independent of u_r . The data consists of a time series of length 100 which is generated by system (6). To obtain a time series in l_2 , the observation (u, y) is taken to be zero outside the observation interval.

Of course, this simple example could be solved by brute force as a nonlinear parameter optimization problem, disregarding any system theoretic interpretation of the problem. In more complicated cases, however, this becomes hardly feasible. Therefore we will follow a system theoretic approach, that also gives more insight in the problem.

The GTLS results in the following sections will be compared with those obtained by three other methods, namely regression, the “output error” method, and the “local total least squares” method. Here we mention that the procedures for GTLS and local total least squares have been implemented in Matlab and that for the regression and output error method we used, respectively, the procedures ARX and OE of the System Identification Toolbox. The regression model B_{regr} is obtained by regressing $y(t)$ on $y(t-1)$, $u(t)$ and $u(t-1)$. For the observation w this gives

$$B_{\text{regr}} = \{(u, y) \in l_2^2; y(t) = 0.10y(t-1) + 1.42u(t) - 0.78u(t-1)\}. \quad (7)$$

This equation yields optimal one-step ahead predictions for $y(t)$, given $y(t-1)$, $u(t)$, and $u(t-1)$. According to our terminology, we would call this method “local ordinary least squares.” By ordinary we mean that only one of the components of w is approximated and by local that only

the first step ahead predictions of difference equations are taken into account. In the GTLS scheme we approximate both components and take full account of the global, higher order forward and backward implications of difference equations.

The output error model B_{oe} is the system with the property that for the given input u , the corresponding system output \hat{y}_{oe} is as close as possible to the observed output y . In fact, this method has some similarity to GTLS, the difference being that the input is kept fixed and only the output is approximated. In our terminology it is a “global ordinary least squares” method. The estimated output error model is

$$B_{\text{oe}} = \{(u, y) \in l_2^2; y(t) = 0.68y(t-1) + 1.15u(t) - 1.51u(t-1)\}. \quad (8)$$

Finally we use a simple modification of the static total least squares method of Section II to determine a first-order model, as follows. Define the block Hankel matrix $H \in l_2^4$ by $H(t) = \begin{bmatrix} w(t-1) \\ w(t) \end{bmatrix}$, then static equations for H correspond to first-order equations for the observation w . The optimal static equation for H is obtained by applying Proposition 2.2 with $q = 4$ and $m = 3$. Clearly, the quality of the corresponding first-order model for w is evaluated only locally. By this we mean that, for example, the second-order restrictions on $(w(t-2), w(t-1), w(t))^T$ implied by the model are not taken into account, and the same holds true for higher order restrictions. Therefore we call this the “local total least squares” model. For the data of this example this gives

$$B_{\text{tls}} = \{(u, y) \in l_2^2; y(t) = 0.54y(t-1) + 1.87u(t) - 1.82u(t-1)\}. \quad (9)$$

This example will be continued in the next sections. \diamond

IV. STATE REPRESENTATIONS

One of the crucial questions in our modeling theory is how to calculate the misfit of a system with respect to a given observation, cf. Definition 3.3. Obviously this requires a numerical representation of the system. In Section V we develop a representation that is extremely useful for this purpose, namely the isometric state representations. They also play a central role in the construction of optimal models, as discussed in Section VII. We now first introduce general state representations, which will be abbreviated as SR. Let σ denote the shift operator, defined as $\sigma x(t) := x(t+1)$.

Definition 4.1 (State Representation): A state representation (A, B, C, D) of an l_2 -system B is a description of the form

$$B = \{w \in l_2^q; \exists x \in l_2^n, v \in l_2^m \text{ such that} \\ \sigma x = Ax + Bv \text{ and } w = Cx + Dv\}$$

with $A \in \mathbf{R}^{n \times n}$, $B \in \mathbf{R}^{n \times m}$, $C \in \mathbf{R}^{q \times n}$, $D \in \mathbf{R}^{q \times m}$ and $n, m \in \mathbf{N}$.

Here v is an auxiliary input, x is a state trajectory and w a system trajectory, m denotes the number of auxiliary inputs and n the number of state variables. The system defined by

this representation is denoted as $B(A, B, C, D)$. We mention that the condition that x and v are square summable involves no loss of generality, as it can be shown that for every $(x, v): \mathbf{Z} \rightarrow \mathbf{R}^n \times \mathbf{R}^m$ generating $w \in l_2^q$ there also exists $(x', v') \in l_2^n \times l_2^m$ generating the same time series w . Systems in $\mathbf{B}^{q,m,n}$, i.e., with rank at most m and degree at most n , are precisely those systems that admit an SR with m auxiliary inputs and n states, cf. Proposition A.4.

Representations are called equivalent if they describe the same system. They are called minimal if the number of state variables and the number of auxiliary inputs are both minimal. These two quantities can be minimized simultaneously, so for every system there indeed exists a minimal SR. From a given SR we obtain equivalent ones as follows.

Proposition 4.2 (Equivalent State Representations): Let be given a state representation (A, B, C, D) of an l_2 -system \mathcal{B} . Then for all invertible $S \in \mathbf{R}^{n \times n}$, invertible $R \in \mathbf{R}^{m \times m}$, and $F \in \mathbf{R}^{m \times n}$, $(S(A + BF)S^{-1}, SBR, (C + DF)S^{-1}, DR)$ is also a state representation of \mathcal{B} . Moreover, if (A, B, C, D) is minimal, then all minimal state representations for \mathcal{B} are obtained in this way.

Proof: By definition, $w \in l_2^q$ is contained in $B(A, B, C, D)$ if and only if there exist an v and x such that $\sigma x = Ax + Bv$ and $w = Cx + Dv$. These equations are equivalent to $\sigma(Sx) = S(A + BF)S^{-1}(Sx) + SBR(R^{-1}(v - Fx))$ and $w = (C + DF)S^{-1}(Sx) + DR(R^{-1}(v - Fx))$, provided that S and R are invertible. This shows the equivalence of the representations. For a proof of the fact that for minimal representations all equivalent representations are obtained in this way we refer to [5, Corollary IV.3-4]. \square

Observe that minimal SR's for a given l_2 -system are highly nonunique. The choice of basis in the state space, corresponding to S , is a well-known nonuniqueness of state-space representations. In our framework the auxiliary input v is merely a tool to describe the system behavior and need not have additional external significance. This allows for a basis transformation for the auxiliary input, corresponding to R . Further the behavior is invariant under a static state feedback F to this auxiliary input. This is in contrast to the common notion of feedback to the actual input of the system, which would affect the set of compatible input-output pairs. As a consequence, in our framework the spectrum of the A -matrix is not an intrinsic property of a system. In the next section we exploit this nonuniqueness to obtain SR's with convenient properties for computing the misfit of a model with respect to data.

Minimality of a representation can be expressed in terms of rank conditions on the matrices (A, B, C, D) , as follows.

Proposition 4.3 (Minimal State Representations for l_2 -Systems): A state representation (A, B, C, D) is minimal if and only if:

- 1) (A, B) is controllable
- 2) $\forall F \in \mathbf{R}^{m \times n}$, $(A + BF, C + DF)$ is observable
- 3) $\ker D = 0$.

Proof: See Appendix B.

By eliminating the state, SR's induce an image representation of l_2 -systems. For the exposition in the sequel,

representations with A asymptotically stable are most relevant. Every l_2 -system allows for such a representation, which follows from Propositions 4.2 and 4.3.1.

Proposition 4.4 (Image Representations): Let \mathcal{B} be an l_2^q -system in $\mathbf{B}^{q,m,n}$, with state representation (A, B, C, D) , where A is asymptotically stable. Let $G: l_2^m \rightarrow l_2^q$ be defined as $G = C(\sigma I - A)^{-1}B + D$, i.e., $(Gv)(t) = Dv(t) + \sum_{k=1}^{\infty} CA^{k-1}Bv(t-k)$. Then $\mathcal{B} = \text{im } G$.

Proof: G is well defined, as A is asymptotically stable. It is easily verified that the equations $\sigma x = Ax + Bv$ and $w = Cx + Dv$ imply that $Gv = w$. \square

Leading Example (Continued): Consider the l_2 -system corresponding to (5), $\mathcal{B}_{\text{ex}} = \{w \in l_2^2; w = (u, y), \text{ with } y(t) = 2/3y(t-1) + 2u(t) - 2u(t-1)\}$. This can be written in input/state/output form as follows

$$x(t+1) = 2/3x(t) + u(t); y(t) = -2/3x(t) + 2u(t). \quad (10)$$

From this it is easy to obtain an SR by taking the auxiliary input v to be equal to u , which gives

$$\mathcal{B}_{\text{ex}} = \mathcal{B}\left(2/3, 1, \begin{bmatrix} 0 \\ -2/3 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}\right). \quad (11)$$

This representation is minimal. The corresponding image representation $G = \sum_{k=0}^{\infty} G_k \sigma^{-k}$ has coefficients $G_0 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$, and $G_k = \begin{bmatrix} 0 \\ -(2/3)^k \end{bmatrix}$. An SR equivalent to (11) is given by $\left(1, 1/2, \begin{bmatrix} 1/3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1 \end{bmatrix}\right)$, which follows from Proposition 4.2 by taking $S = 1$, $R = 1/2$, and $F = 1/3$. Note that in this representation the auxiliary input is given by y . This also illustrates that the eigenvalues of A are not intrinsic for the system \mathcal{B} . \diamond

V. ISOMETRIC STATE REPRESENTATIONS

In this section we define isometric state representations (ISR's), which are defined by a local isometry property involving the state variable.

Definition 5.1 (Isometric State Representation): A state representation (A, B, C, D) is called isometric if for all $x \in \mathbf{R}^n$, $v \in \mathbf{R}^m$, $w \in \mathbf{R}^q$ and $z \in \mathbf{R}^n$ such that $z = Ax + Bv$ and $w = Cx + Dv$ there holds

$$|v|^2 + |x|^2 = |w|^2 + |z|^2. \quad (12)$$

Equivalently

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^T \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ 0 & I_m \end{pmatrix}. \quad (13)$$

Minimal ISR's can be constructed from arbitrary minimal SR's as follows.

Proposition 5.2 (Construction of ISR): Let (A, B, C, D) be a minimal state representation, and let $K \in \mathbf{R}^{n \times n}$ be the unique symmetric positive definite solution of the algebraic Riccati equation

$$K = A^T K A - (B^T K A + D^T C)^T (B^T K B + D^T D)^{-1} \cdot (B^T K A + D^T C) + C^T C. \quad (14)$$

Let the matrices $S \in \mathbf{R}^{n \times n}$, $F \in \mathbf{R}^{m \times n}$ and $R \in \mathbf{R}^{m \times m}$ be solutions of the equations

$$S^T S = K \quad (15)$$

$$RR^T = (B^T K B + D^T D)^{-1} \quad (16)$$

$$F = -(B^T K B + D^T D)^{-1} (B^T K A + D^T C). \quad (17)$$

Then $(S(A + BF)S^{-1}, SBR, (C + DF)S^{-1}, DR)$ is an equivalent isometric minimal state representation.

Proof: See Appendix B.

The following proposition gives a necessary minimality condition for ISR's and states that they are unique modulo unitary transformations.

Proposition 5.3 (Minimal ISR):

- 1) If (A, B, C, D) is an ISR, then A is stable. If the representation is minimal then A is asymptotically stable.
- 2) Two minimal ISR's (A, B, C, D) and (A', B', C', D') are equivalent if and only if there exist unitary matrices U and V such that $(A', B', C', D') = (UAU^T, UB, CV, DV)$.

Proof: See Appendix B.

ISR's induce a description of l_2 -systems as the image of an isometric operator. This is made explicit in the following proposition.

Proposition 5.4 (Isometric Image Representations): Let (A, B, C, D) be a minimal isometric state representation of an l_2 -system \mathcal{B} . Then the corresponding image operator $G: l_2^m \rightarrow l_2^q$ defined by $G = C(\sigma I - A)^{-1}B + D$ is isometric, i.e., $\|Gv\| = \|v\|$ for all $v \in l_2^m$.

Proof: If $w = Gv$ then $w(t) = Cx(t) + Dv(t)$ for $x = (\sigma I - A)^{-1}Bv$, so that $\sigma x = Ax + Bv$. Summation of (12) over $t \in \mathbf{Z}$ yields $\|v\|^2 + \|x\|^2 = \|w\|^2 + \|\sigma x\|^2$. Clearly $\|x\| = \|\sigma x\|$, from which the result follows. \square

Summarizing, every l_2 -system can be represented as the image of an isometric $q \times m$ transfer function with an ISR as its realization. In the literature an isometric operator G is sometimes called lossless, and if it is in addition stable it is called inner. So an ISR is a realization of an inner transfer function that displays the isometry in a local way, in terms of the state variables.

Leading Example (Continued): We apply the construction of Proposition 5.2 to the SR (11) of l_2 -system \mathcal{B}_{ex} . This yields $K = 4/9$, $S = 2/3$, $R = 3/7$, and $F = 4/21$, resulting in the isometric representation

$$\sigma x = 6/7x + 2/7v; w = \begin{bmatrix} 2/7 \\ -3/7 \end{bmatrix} x + \begin{bmatrix} 3/7 \\ 6/7 \end{bmatrix} v. \quad (18)$$

According to Proposition 5.3, \mathcal{B}_{ex} has a unique minimal ISR, modulo sign changes for the state and for the auxiliary input. \diamond

VI. OPTIMAL APPROXIMATION WITHIN A SYSTEM

In this and the next section we consider the GTLS problem of Definition 3.4. As stated before, this involves a double minimization. In this section we discuss the computation of

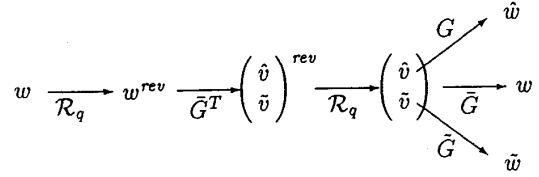


Fig. 2. Projection scheme.

the misfit, and in the next section we treat the problem of selecting an optimal model.

For the computation of the misfit we will use the adjoint G^* of an operator $G: l_2^m \rightarrow l_2^q$. This is defined by the condition $\langle w, Gv \rangle = \langle G^*w, v \rangle$ for all $v \in l_2^m$ and $w \in l_2^q$, where $\langle \cdot, \cdot \rangle$ denotes the inner product on l_2 . It follows that G is isometric if and only if $G^*G = I$. For the image operator $G := C(\sigma I - A)^{-1}B + D$ the adjoint is given by $G^* = B^T(\sigma^{-1}I - A^T)^{-1}C^T + D^T$. This can also be written as $G^* = \mathcal{R}_m G^T \mathcal{R}_q$ where $\mathcal{R}_k: l_2^k \rightarrow l_2^k$ denotes the time reversion operator defined by $\mathcal{R}_k w(t) := w(-t)$ and $G^T := B^T(\sigma I - A^T)^{-1}C^T + D^T$.

Theorem 6.1 (Optimal Approximation Within a System): Let be given an observation $w \in l_2^q$ and an l_2^q -system $\mathcal{B} \in \mathbf{B}^{q,m,n}$. Let $G: l_2^m \rightarrow l_2^q$ be an isometric image representation of \mathcal{B} . Then the optimal approximation $\hat{w} \in \mathcal{B}$ is given by $\hat{w} = GG^*w$, with misfit $d(w, \mathcal{B}) = \|(I_q - GG^*)w\|$.

Proof: This is a well-known result. To be explicit, we should minimize $\|w - Gv\|$ over $v \in l_2^m$. Let $v' = v - G^*w$, then $\|w - Gv\|^2 = \|w - GG^*w - Gv'\|^2 = \|w - GG^*w\|^2 + \|Gv'\|^2$, as $\langle w - GG^*w, Gv' \rangle = \langle G^*w - G^*w, v' \rangle = 0$ where we use that G is isometric so that $G^*G = I_m$. So the minimum is achieved by taking $v' = 0$, and $\hat{w} = GG^*w$. \square

The optimal approximation within a system \mathcal{B} gives rise to a decomposition $w = \hat{w} + \tilde{w}$, with $\hat{w} \in \mathcal{B}$ regular and with \tilde{w} the corresponding approximation error. We will now show that \tilde{w} also exhibits regularity. As \hat{w} is obtained as the projection of w on \mathcal{B} , it follows that $\tilde{w} \in \mathcal{B}^\perp = \{w \in l_2; \langle w, w' \rangle = 0 \text{ for all } w' \in \mathcal{B}\}$. This set is clearly linear and shift-invariant. The following theorem states that it is an l_2 -system.

Theorem 6.2 (Orthogonal Complement): Let \mathcal{B} be an l_2 -system with rank m and degree n .

- 1) The orthogonal complement \mathcal{B}^\perp of \mathcal{B} is an l_2 -system.
- 2) $\mathcal{B} \oplus \mathcal{B}^\perp = l_2^q$.
- 3) \mathcal{B}^\perp has rank $q - m$ and degree n .
- 4) Let (A, B, C, D) be a minimal isometric state representation of \mathcal{B} , and let \tilde{A}, \tilde{B} be such that $\begin{pmatrix} A & B & \tilde{B} \\ C & D & \tilde{D} \end{pmatrix}$ is a unitary matrix. Then $(A, \tilde{B}, C, \tilde{D})$ is a minimal isometric state representation of \mathcal{B}^\perp .
- 5) Let \hat{w} be the optimal approximation in \mathcal{B} of $w \in l_2^q$. Then the approximation error $\tilde{w} := w - \hat{w}$ is the optimal approximation in \mathcal{B}^\perp of w .

Proof: See Appendix B.

We summarize this result in the following projection scheme as shown in Fig. 2. Here G and \bar{G} denote the image representations corresponding to, respectively, (A, B, C, D) and $(A, \tilde{B}, C, \tilde{D})$, and $\bar{G} := [G \ G]$. In the literature \bar{G} is sometimes called a lossless embedding of G . Further \hat{v} and \tilde{v} are

TABLE II
MISFIT OF MODELS

model	\mathcal{B}_{ex}	\mathcal{B}_{oe}	\mathcal{B}_{regr}	\mathcal{B}_{tlls}	\mathcal{B}_{rand}
misfit	5.22	5.62	6.84	5.30	13.53

the auxiliary inputs corresponding to \hat{w} and \tilde{w} , i.e., $\hat{w} = G\hat{v}$ and $\tilde{w} = G\tilde{v}$. According to Theorem 6.1 and 6.2.5, there holds that $\hat{v} = G^*w$ and $\tilde{v} = G^*w$ as shown in Fig. 2.

Leading Example (Continued): We apply the projection algorithm of Theorem 6.1 to determine the optimal approximation $\hat{w} \in \mathcal{B}_{ex}$ of the observation w described in Section III. Let G denote the image operator corresponding to the ISR (18), and let (A, B, C, D) denote the corresponding matrices. First compute $\hat{v} := G^*w$, which is given by the backward state equations $x(t) = A^T x(t+1) + C^T w(t)$; $\hat{v}(t) = B^T x(t+1) + D^T w(t)$. Then $\hat{w} = G\hat{v}$ is given by $\hat{x}(t+1) = A\hat{x}(t) + B\hat{v}(t)$; $\hat{w}(t) = C\hat{x}(t) + D\hat{v}(t)$.

According to Theorem 6.2, \mathcal{B}_{ex}^\perp has ISR $(6/7, 3/7, \begin{bmatrix} 2/7 \\ -3/7 \end{bmatrix}, \begin{bmatrix} -6/7 \\ 2/7 \end{bmatrix})$. This corresponds to the equation

$$y(t) = y(t-1) - 1/3u(t) + 1/2u(t-1). \quad (19)$$

So every time series in l_2^2 that is orthogonal to \mathcal{B}_{ex} satisfies this difference equation, in particular the approximation error $\tilde{w} := w - \hat{w}$. The misfit of \mathcal{B}_{ex} equals $\|\tilde{w}\| = 5.22$, as compared to $\|w\| = 22.72$. This is considerably smaller than the l_2 -norm of the white noise by which the observation was corrupted. Recall that $w = w_r + w_n$, with $w_r \in \mathcal{B}_{ex}$ and w_n white noise. In our example, $\|w_n\| = 7.62$. The optimal approximation is simply obtained by projecting the noise w_n on \mathcal{B}_{ex} , with a resulting decomposition $w_n = \hat{w}_n + \tilde{w}_n$, where $\hat{w}_n \in \mathcal{B}_{ex}$ and $\tilde{w}_n \in \mathcal{B}_{ex}^\perp$. In our case, $\|\hat{w}_n\| = 5.56$ and $\|\tilde{w}_n\| = \|\tilde{w}\| = 5.22$. We also determine projections onto the models \mathcal{B}_{regr} , \mathcal{B}_{oe} , and \mathcal{B}_{tlls} , as defined in Section III, and onto a randomly chosen system $\mathcal{B}_{rand} \in \mathcal{B}^{2,1,1}$. The parameters of an SR of \mathcal{B}_{rand} were obtained by a random sample from the standard normal distribution. The resulting misfits are listed in Table II. It turns out that the local total least squares model is of relatively good quality in this example. This, however, may be completely different in other situations, as we will illustrate by an example in Section IX-B. \diamond

VII. MODEL IMPROVEMENT CONSTRUCTIONS

In this section we discuss the second part of the GTLS problem, namely determining an l_2 -system with minimal misfit with respect to a given observation. Formulated in terms of state representations, this amounts to the following.

Definition 7.1 (GTLS in terms of SR): For given observation $w \in l_2^q$ and tolerated complexity (m, n) , determine a state representation (A, B, C, D) with m auxiliary inputs and n states and an auxiliary input $\hat{v} \in l_2^m$ such that for $\hat{w}(t) := (C(\sigma I - A)^{-1}B + D)\hat{v}$, the error $\|w - \hat{w}\|$ is minimal.

We follow an iterative approach for this nonlinear problem. In each step we keep some parameters fixed, such that the resulting subproblem becomes sufficiently simple. For instance, for fixed (A, B, C, D) the resulting problem in \hat{v} is solved

by the projection scheme discussed in the foregoing section. We consider the following subproblems.

Problem 1—Optimal C and D: For given A, B and \hat{v} , solve the GTLS problem for C and D .

Problem 2—Optimal B and D: For given A, C and \hat{v} , solve the GTLS problem for B and D .

Problem 3—Optimal B, D and \hat{v} in an ISR: For given A and C with $A^T A + C^T C = I_n$, solve the GTLS problem for B, D , and \hat{v} , under the restriction that (A, B, C, D) is isometric.

In the next theorem we give constructive solutions of these problems.

Theorem 7.2 (Model Improvement Constructions):

- 1) Construction 1 (Projection of the Approximation Error): Let \hat{x} be defined by $\sigma\hat{x} = A\hat{x} + B\hat{v}$, and let $\mathcal{E} := \{\hat{w} \in l_2^q; \exists C \in \mathbb{R}^{q \times n}, D \in \mathbb{R}^{q \times m} \text{ such that } \hat{w} = C\hat{x} + D\hat{v}\}$. Let $P\hat{x} + Q\hat{v}$ denote the orthogonal projection of w onto \mathcal{E} , then $\mathcal{B}_1 := \mathcal{B}(A, B, P, Q)$ solves problem 1.
- 2) Construction 2 (Dual version of construction 1): Let $\mathcal{F} := \{\hat{w} \in l_2^q; \exists B \in \mathbb{R}^{n \times m}, D \in \mathbb{R}^{q \times m} \text{ such that } \hat{w} = (C(\sigma I - A)^{-1}B + D)\hat{v}\}$. Let the orthogonal projection of w onto \mathcal{F} be given by $(C(\sigma I - A)^{-1}P + Q)\hat{v}$, then $\mathcal{B}_2 := \mathcal{B}(A, P, C, Q)$ solves problem 2.
- 3) Construction 3 (SVD on auxiliary inputs): Let $\bar{B} \in \mathbb{R}^{n \times q}$, $\bar{D} \in \mathbb{R}^{q \times q}$ be such that $\begin{pmatrix} \bar{B} \\ \bar{D} \end{pmatrix}$ is a unitary matrix. Let $\bar{v} := (\bar{B}^T(\sigma^{-1}I - A^T)^{-1}C^T + \bar{D}^T)w$ have SVD $\bar{v} = \sum_{i=1}^q \lambda_i u_i z_i$, and let $U_m := [u_1, \dots, u_m]$. Then $\mathcal{B}_3 := \mathcal{B}(A, \bar{B}U_m, C, \bar{D}U_m)$ solves problem 3.

Proof: Parts 1 and 2 follow immediately from the definitions of \mathcal{E} and \mathcal{F} . For part 3, observe that (A, B, C, D) is an ISR if and only if there exist \tilde{B}, \tilde{D} , and a unitary V such that $[B \ \tilde{B}] = \bar{B}V$ and $[D \ \tilde{D}] = \bar{D}V$. For this representation the projection scheme gives $\begin{pmatrix} \hat{v} \\ \tilde{v} \end{pmatrix} = V^T \bar{v}$, with misfit $\|\bar{v}\|$. By taking $V = [u_1, \dots, u_q]$, this misfit is determined by the $q-m$ smallest singular values of \bar{v} , which is minimal. \square

We use these results in an iterative algorithm for the GTLS problem. At each step we use one of the three constructions to improve the model. The resulting model parameters are transformed to ISR, which also involves an update of the A -matrix. The projection scheme is then applied to update \hat{v} .

Proposition 7.3: The above method leads to a sequence of models with monotonically decreasing misfit.

Proof: This is immediately evident from Theorem 7.2. \square

So the algorithm leads to a convergent sequence of misfits. In the next section we show that, in the limit, the corresponding models are stationary points with respect to the GTLS criterion.

Leading Example (Continued): To illustrate the foregoing, we consider the data $w = (u, y)$ described in Section III (see (5) and (6)). To investigate the effect of the choice of an initial model, we apply the model improvement constructions to the models \mathcal{B}_{ex} , \mathcal{B}_{oe} , \mathcal{B}_{regr} , \mathcal{B}_{tlls} , and \mathcal{B}_{rand} as described in Sections III and VI. The results are in Table III. The first row shows the initial misfits, cf. Table II. The next three rows contain the misfits of the models obtained by applying each of the constructions separately and only once. This shows that

TABLE III
MODEL IMPROVEMENTS

model	B_{ex}	B_{oe}	B_{reg}	B_{tls}	B_{rand}
misfit	5.22	5.62	6.84	5.30	13.53
construction 1	5.14	5.12	6.10	5.23	6.21
construction 2	5.14	5.13	6.13	5.30	6.26
construction 3	5.21	5.15	6.78	5.30	6.59
limit	5.11	5.11	5.11	5.11	5.11

TABLE IV
ORDER SELECTION

order	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
misfit	6.71	5.11	5.02	4.91	4.91

each individual construction can give a significant decrease of the misfit. The last row shows the misfit resulting from applying these constructions iteratively until convergence.

For each initial model convergence occurred after about 20 iterations. The limiting model is the same in all four cases, which suggests that it is optimal. It is given by

$$B_{gtls} = \{(u, y) \in l_2^2; y(t) = 0.67y(t-1) + 1.85u(t) - 1.96u(t-1)\}. \quad (20)$$

The parameters of this system are relatively close to those of the data generating system, cf. (5).

Next, we investigate whether the model order can be deduced from the data. For this purpose we compare in Table IV the optimal misfits for models of various degree. The misfit of the optimal static model is given by the smallest singular value of w . This clearly motivates the choice of a first-order model. It is significantly better than the static model, and an increase of the order gives only small improvements. \diamond

VIII. OPTIMALITY CONDITIONS

From the model improvement constructions in Section VII we can derive necessary conditions for optimality, as for an optimal model the constructions can give no improvement. We express the optimality conditions in terms of empirical covariances. For two sequences $a \in l_2^k$ and $b \in l_2^l$ this is defined as $\text{cov}(a, b) := \sum_{t=-\infty}^{\infty} a(t)b(t)^T \in \mathbb{R}^{k \times l}$. Further, by $\text{cov}([a_1, a_2], [b_1, b_2])$, we denote the covariance matrix of the combined trajectories $[a_1^T a_2^T]^T$ and $[b_1^T b_2^T]^T$.

Theorem 8.1 (l_2 -optimality Conditions): Let B denote a GTLS model for an observation $w \in l_2^2$. Let $\hat{w} \in B$ denote the optimal approximation of w , and $\tilde{w} \in B^\perp$ the corresponding approximation error. Let \hat{x}, \hat{v} denote, respectively, the state and auxiliary input corresponding to \hat{w} in a minimal state representation of B , and let \tilde{x}, \tilde{v} be defined analogously for \tilde{w} . Then the following equivalent conditions hold:

- 1) $\text{cov}(\hat{v}, \tilde{v}) = 0$, $\text{cov}(\hat{v}, \tilde{x}) = 0$ and $\text{cov}(\tilde{x}, \tilde{v}) = 0$;
- 2) $\text{cov}([\hat{v}, \hat{x}], [\tilde{v}, \tilde{x}]) = 0$;
- 3) $\text{cov}([\hat{v}, \hat{x}, \hat{w}, \sigma\hat{x}], [\tilde{v}, \tilde{x}, \tilde{w}, \sigma\tilde{x}]) = 0$.

Proof: See Appendix B.

In practice, to evaluate how far these conditions are satisfied it may be useful to consider the empirical correlations, i.e., the covariances scaled by the magnitude of the variables.

Next we investigate how far these conditions are sufficient for optimality. It is not difficult to check that the number of free parameters in (A, B, C, D) , modulo the equivalence of Proposition 4.2, is given by $nq + m(q - m)$. This is precisely the number of equations in Theorem 8.1.1. In fact, these conditions characterize the stationary points with respect to the GTLS criterion. We call a system B a stationary point for an observation w if all the derivatives of the GTLS misfit $d(w, B(A, B, C, D))$ with respect to the system parameters are zero for a minimal SR of B . As all its minimal SR's are linearly related this is equivalent to the condition that all minimal SR's of B are stationary points.

Theorem 8.2: An l_2 -system B satisfies the optimality conditions of Theorem 8.1 if and only if B is a stationary point of the GTLS criterion.

Proof: See Appendix B.

This shows that the GTLS algorithm can only converge to stationary points. This does not, however, establish convergence of the systems. For what it is worth, we mention that we never encountered convergence problems in any of our simulations. A thorough discussion of the convergence properties of the algorithm falls beyond the scope of this paper.

The foregoing results can also be used to analyze whether a proposed system B is close to optimality. This is, for example, relevant in the formulation of stopping criteria for the iterative algorithm of Section VII. Probably the most convincing way to evaluate optimality is to consider the distance between B and a GTLS model B^* , as defined in Definition 3.4. This is in general not feasible, however, as it would require the knowledge of B^* . Instead of asking how far the system should be changed to become optimal for the observed data w , we will consider the question of how far these data should be changed to make the given system optimal. For pragmatic reasons we consider the distance to the nearest stationary point, defined as

$$\min\{\|\bar{w}\|; B \text{ is stationary for } w - \bar{w}\}. \quad (21)$$

Because it seems difficult to evaluate this distance exactly, we present an upper bound that is relatively easy to compute. This upper bound is obtained by allowing only adjustments of the data that belong to B^\perp , so that the optimal approximation of the data within B is not affected. This leads to the following definition of the optimality margin.

Definition 8.3 (Optimality Margin): The optimality margin of a system B with respect to an observation w is defined as

$$\min\{\|\bar{w}\|; \bar{w} \in B^\perp \text{ and } B \text{ is stationary for } w - \bar{w}\}. \quad (22)$$

The following result shows that the computation of the optimality margin is indeed relatively easy.

Proposition 8.4 (Optimality Margin): Let \hat{w} denote the optimal approximation of w in B and let $\tilde{w} := w - \hat{w}$ denote the corresponding approximation error. Further define $\mathcal{Z} := \{z \in B^\perp; B \text{ is stationary for } \hat{w} + z\}$. Then \mathcal{Z} is a linear space, and the optimality margin is given by $\|\tilde{w} - \tilde{w}'\|$, where \tilde{w}' is the orthogonal projection of \tilde{w} on \mathcal{Z} .

TABLE V
OPTIMALITY MARGINS

model	B_{gtls}	B_{ex}	B_{oe}	B_{reg}	B_{ltls}	B_{rand}
optimality margin	$4.58 * 10^{-6}$	1.00	2.29	3.32	1.23	11.19

Proof: See Appendix B.

These results can also be used to determine a lower bound for the achievable misfit. This indicates the quality of a proposed model relative to the optimal one. For this purpose we will assume that a proposed model B is not only stationary for the adjusted data $w - \bar{w}$, but even globally optimal.

Proposition 8.5 (Bounds for the Minimal Misfit): Let e be given an observation w , and let e^* be the minimally achievable misfit under a certain complexity constraint. Further let B be a GTLS model of tolerated complexity for adjusted data $w - \bar{w}$ with misfit $e := d(w, B)$. Then there holds

$$e - 2\|\bar{w}\| \leq e^*. \quad (23)$$

Proof: Let B^* denote a GTLS model for the original data w , so that $d(w, B^*) = e^*$. Then the upper bound follows from the optimality of B^* . For the lower bound we use the properties of the misfit that $d(w, B) \leq \|w\|$ and $d(w_1 + w_2, B) \leq d(w_1, B) + d(w_2, B)$, so that $e = d(w, B) \leq d(\bar{w}, B) + d(w - \bar{w}, B) \leq \|\bar{w}\| + d(w - \bar{w}, B^*) \leq \|\bar{w}\| + d(w, B^*) + d(\bar{w}, B^*) \leq 2\|\bar{w}\| + e^*$. Here we have used the optimality of B for $w - \bar{w}$ in the second inequality. \square

This shows that for models with a small optimality margin the corresponding misfit is nearly optimal.

Leading Example (Continued): The covariances of Theorem 8.1 give a first indication of the optimality of a model. To make this scale invariant we consider the correlations in an ISR. For the nominal model B_{ex} they are around 0.3, for the regression model B_{reg} around 0.35, while for the randomly chosen model B_{rand} correlations of 0.8 occur. For the model B_{gtls} in (20) the correlations are approximately zero, below 10^{-5} .

From the optimality margins we obtain more precise information about the optimality of the systems. They are listed in Table V. This shows that it requires only a change of the observation of the order $\|\bar{w}\| \approx 10^{-5}$ to make B_{gtls} a stationary point, cf. (22). Now assume that B_{gtls} is globally optimal for $w - \bar{w}$, which is reasonable assumption. The evaluation of the bounds in Proposition 8.5 for the data in this example with $e = 5.1095$ and $2\|\bar{w}\| \approx 10^{-5}$ shows that the optimal misfit $e^* \approx 5.1095$ is determined within an accuracy of 10^{-5} . This also shows that B_{gtls} is optimal within this accuracy level.

IX. SIMULATION EXPERIMENTS

We illustrate the use of the GTLS algorithm by three simulation experiments. The first example concerns model reduction, i.e., the approximation of a system by one of lower complexity. We use weighted l_2 -norms to determine the l_2 -optimal approximation of a systems impulse response. In the second example we show that the algorithm can handle noncausal systems without any additional difficulty. As a final example we identify a system with multiple outputs, described

TABLE VI
PROPERTIES OF REDUCED MODELS

model	B_{gtls}	B_{bal}	B_{hank}
misfit with respect to w	0.32	0.36	0.38
error in impulse response	0.54	0.46	0.57
Hankel norm distance	0.93	0.80	0.74

TABLE VII
EFFECT OF SCALING

scaling factor	$\alpha = 1$	$\alpha = 10$	$\alpha = 100$
$\ \cdot\ _\alpha$ -misfit with respect to w	0.3204	0.4457	0.4476
error in impulse response	0.5407	0.4477	0.4476

by more than one equation, and we discuss the choice of the model complexity.

A. l_2 -Model Reduction

The algorithm of Section VII can be applied to arbitrary time series in l_2 . Here we analyze its performance for very special data, a system impulse response. The aim is to reduce the dimension of the state space in such a way that the error in the impulse response is as small as possible (cf. [9] and the references therein). We compare the results of our algorithm with those obtained by balanced reduction and optimal Hankel norm approximation that have been developed especially for model reduction (see [2] and [7]). We consider the single-input, single-output system B with poles in $\pm 0.9i$ and $-0.7 \pm 0.6i$, so $B = \{[u \ y]^T \in l_2^2; y(t) = 0.5u(t) - 1.4y(t-1) - 1.66y(t-2) - 1.13y(t-3) - 0.69y(t-4)\}$. This system has complexity (1, 4), and we consider reduction to complexity (1, 2). The observation $w \in B$ consists of two components u and y , where u is a unit pulse at time $t = 0$ and y is the corresponding response.

We apply the GTLS algorithm, starting in a randomly chosen model. When the decrease in the misfit has become sufficiently small, below 10^{-10} , the iterations are stopped. This occurs after a few hundred iterations. The final model B_{gtls} is compared in Table VI with the balanced reduction B_{bal} and the Hankel norm reduction B_{hank} .

The l_2 -error in the impulse response in B_{gtls} is somewhat larger than that in B_{bal} and B_{hank} . If one is interested in this response then one should prevent an approximation of the input, so that an optimal approximation of the output becomes the criterion. This is achieved by taking the norm $\|\hat{w}\|_\alpha^2 := \alpha^2 \|\hat{u}\|^2 + \|\hat{y}\|^2$ with α sufficiently large. The effect of increasing α is given in Table VII. This shows that for large α the method determines better approximations of the impulse response.

This also gives bounds for the minimally achievable l_2 -error, which we denote by e^* . Let B_α be the GTLS model for $\|\cdot\|_\alpha$, and let y_α be the impulse response of B_α ; then it is easily checked that $d(w, B_\alpha) \leq e^* \leq \|y - y_\alpha\|$. By increasing α we can obtain an arbitrarily accurate estimate of e^* . This gives an iterative solution method for the l_2 -optimal impulse response approximation problem. For $\alpha = 100$ we obtain $e^* = 0.4476$; see Table VII. The corresponding model

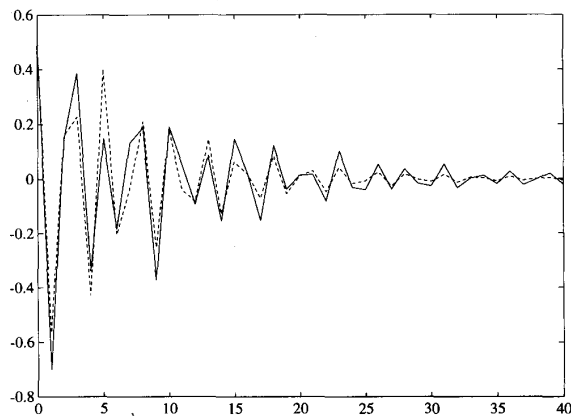


Fig. 3. The impulse response (solid line) and its l_2 -optimal approximation of second order (dashed line).

is given by the equation $y(t) = -1.35y(t-1) - 0.78y(t-2) + 0.50u(t) + 0.11u(t-1) - 0.21u(t-2)$. The impulse response is depicted in Fig. 3.

B. Noncausal Systems

We consider the following noncausal system, the “Mexican hat,”

$$w_2(t) = -\frac{d^2}{dt^2} \left\{ \int_{-\infty}^{\infty} \varphi(x) w_1(t-x) dx \right\} \quad (24)$$

where φ is the standard normal density $\varphi(x) = (2\pi)^{-1/2} e^{-\frac{1}{2}x^2}$. In the simulations we consider a discrete-time version $w_2(t) = \sum_{j=-N}^N G_j w_1(t-j)$ with $N = 40$ and time steps of size 0.2. Note that w_2 is not a causal output, as the transfer function from w_1 to w_2 is not proper.

First we apply our procedure to the impulse response observation, i.e., w_1 is a unit pulse at time $t = 0$ and w_2 is the corresponding response; see Fig. 4. The misfits of the optimal models of orders 2, 4, and 6 are given in Table VIII. They are compared with the optimal Hankel norm approximations of orders 2, 4, and 6. These are obtained by approximations of orders 1, 2, and 3 of the causal part of the impulse response and using the symmetry of the Mexican hat to estimate the anticausal part. Analogously we determined approximations by balanced reduction. In Table VIII we also list the error in the impulse response of these models, i.e., the l_2 -distance between the systems impulse response and the Mexican hat w_2 . This error should be compared with the magnitude of the response, given by $\|w_2\| = 0.35$.

Hankel norm reduction and especially balancing give rather good results. They can only be used, however, when a causal impulse response is available. The GTLS method makes no use of the symmetry of the observed signals, but this property is preserved well in the identified models. This is illustrated in Figs. 5 and 6, which contain the optimal approximations of orders 2 and 4.

We also apply the GTLS method to data w_n consisting of two noisy steps for the input and the corresponding system output. These data and the optimal approximation of order 4

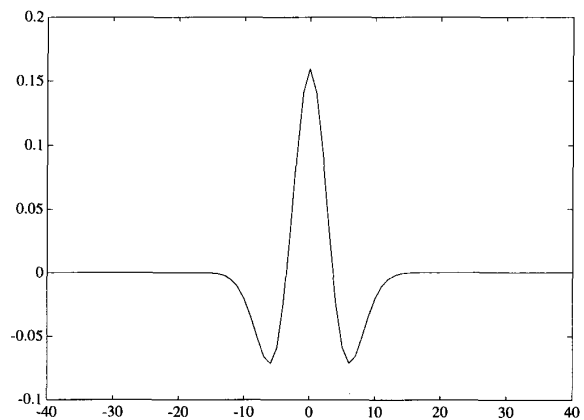


Fig. 4. The Mexican hat.

TABLE VIII
MODEL QUALITY

model		B_{gtls}	B_{bal}	B_{hank}
n=2	misfit	0.17	0.18	0.32
	error in impulse response	0.21	0.20	0.34
n=4	misfit	0.046	0.047	0.052
	error in impulse response	0.051	0.050	0.055
n=6	misfit	0.0070	0.0071	0.0078
	error in impulse response	0.0076	0.0075	0.0081

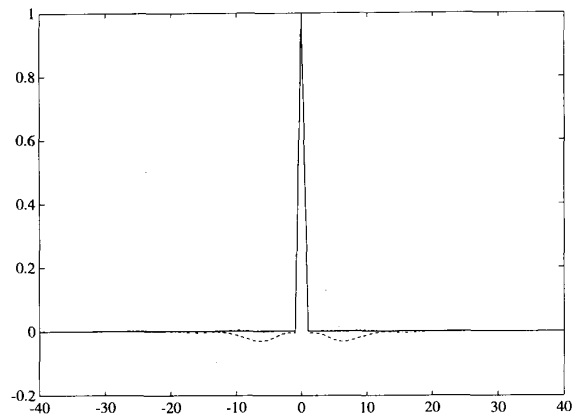


Fig. 5. The pulse w_1 (solid line), the first component of the l_2 -optimal approximation of second order (dashed line) and fourth order (dash-dotted line).

are given in Fig. 7. We should mention that the approximation error in the input is so small that it is nearly invisible in this figure.

The misfit of this model and the error in its impulse response are listed in the first column of Table IX. In view of the results for $n = 4$ in Table VIII, this shows that the identified model is a rather accurate approximation of the Mexican hat. Depending on the choice of an initial model, it typically takes a few hundred iterations to obtain convergence, and sometimes convergence to a local optimum occurred.

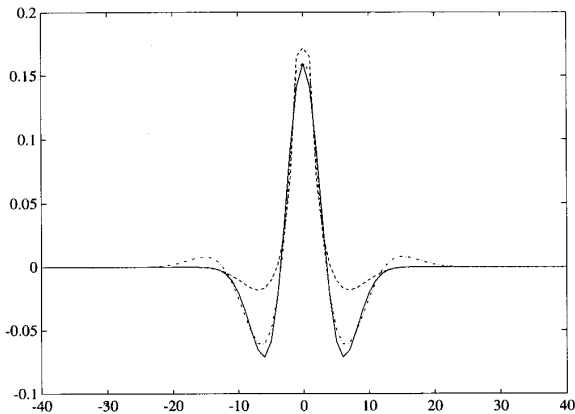


Fig. 6. The Mexican hat w_2 (solid line), the second component of the l_2 -optimal approximation of second order (dashed line) and fourth order (dash-dotted line).

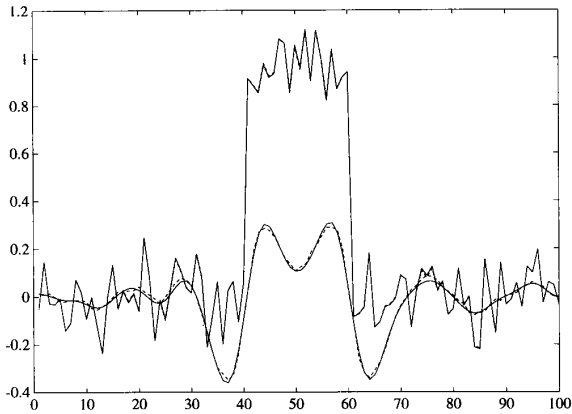


Fig. 7. Noisy step measurement w_n (solid lines) and its l_2 -optimal approximation of fourth order (dashed lines).

TABLE IX
LOCAL VS. GLOBAL METHOD

method	Global TLS	Local TLS
misfit with respect to w_n	0.11	1.17
error in impulse response	0.066	0.522

Finally, as mentioned at the end of Section VI, we will once more consider the local total least squares method described in Section III. The results for the fourth order model are given in the second column in Table IX. This clearly shows that, perhaps not surprisingly, the local method gives poor results with respect to the global total least squares criterion. For example, the error in the impulse response of the local model is even larger than the Mexican hat itself, which has norm $\|w_2\| = 0.35$.

C. A System with Multiple Outputs

In this experiment we consider a system with multiple outputs, so that a single difference equation does not suffice to describe the system. For simplicity we consider a system

TABLE X
SELECTION OF RANK AND DEGREE

model	$n = 0$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
$m = 1$	3.33	1.45	0.93	0.89	0.88
$m = 2$	1.30	0.61	0.56	0.52	0.46

with one input and two outputs. The data are generated as $w = w' + e$, where $w' \in \mathcal{B}^{3,1,2}$ satisfies the equations

$$\begin{aligned} w'_2(t) &= w'_2(t-1) + w'_1(t) \\ w'_3(t) &= w'_2(t) + w'_1(t-1). \end{aligned} \quad (25)$$

For w_1 we take white noise with unit variance, and for e a three-dimensional white noise process with independent components and variance 0.01. The observation interval has length 50. Outside this interval we define $w(t) = 0$. The GTLS model of rank one and degree two is

$$\begin{aligned} w_2(t) &= 0.95w_2(t-1) - 0.04w_2(t-2) + 1.08w_1(t) \\ &\quad + 0.07w_1(t-1) + 0.15w_1(t-2) \\ w_3(t) &= 0.95w_3(t-1) - 0.04w_3(t-2) + 1.12w_1(t) \\ &\quad + 1.08w_1(t-1) - 0.89w_1(t-2). \end{aligned} \quad (26)$$

Transforming model equation (25) to the form

$$\begin{aligned} w'_2(t) - w'_2(t-1) &= w'_1(t) \\ w'_3(t) - w'_3(t-1) &= w'_1(t) + w'_1(t-1) - w'_1(t-2) \end{aligned} \quad (27)$$

shows that the original model equations are estimated rather accurately. We compare the misfits of GTLS models of various complexity in Table X.

For rank one, the misfit hardly decreases for orders above two. This could be expected, as the regular part of the data belongs to a system of order two. For rank two the results suggest to take the order one. Comparing the complexities $(m, n) = (1, 2)$ and $(2, 1)$, the first one of course leads to a larger misfit, as it imposes more restrictions. The misfit is still relatively small, however, when compared to the norm of the data $\|w\| = 6.10$.

X. CONCLUSION

In this paper we investigated the modeling of vector time series by means of difference equations, using the global total least squares criterion. Distinctive features of our approach are that no decomposition into inputs and outputs is required and that the criterion measures the global misfit in a nonparametric way. The misfit of a given system is evaluated by a dynamic projection algorithm formulated in terms of isometric state representations. We developed an iterative algorithm for constructing optimal models and gave a characterization of stationary points of the GTLS criterion. The method was illustrated by some simulation experiments.

The results of this paper can be extended in several directions, e.g., time series on a finite time interval, time varying systems, and time-varying norms for the misfit. Further research will be concerned with statistical properties and the development of faster algorithms and recursive methods.

APPENDIX A

l_2 -SYSTEMS

By l_2^q we denote the set of q -dimensional square summable time series over time axis \mathbf{Z} , i.e., $l_2^q = \{w: \mathbf{Z} \rightarrow \mathbf{R}^q; \sum_{t=-\infty}^{\infty} w(t)^T w(t) < \infty\}$. We define l_2 -systems as follows. By σ we denote the time shift defined by $(\sigma w)(t) := w(t+1)$, $t \in \mathbf{Z}$.

Definition A.1 (l_2 -systems): An l_2 -system is a subset \mathcal{B} of l_2^q that is:

- 1) linear, i.e., for $w, w' \in \mathcal{B}$ and $\alpha, \beta \in \mathbf{R}$ there holds $\alpha w + \beta w' \in \mathcal{B}$;
- 2) shift-invariant, i.e., $\sigma \mathcal{B} = \mathcal{B}$;
- 3) complete, i.e., if $w \in l_2^q$ satisfies $w|_T \in \mathcal{B}|_T$ for all finite $T \subset \mathbf{Z}$, then $w \in \mathcal{B}$.

This class of systems can be represented by linear, time-invariant difference equations of finite lag. Let $\mathbf{R}^{p \times q}[s]$ denote the set of $p \times q$ matrices with polynomial entries. A set of p difference equations corresponds to a polynomial matrix in the shift σ , i.e., $R(\sigma) \in \mathbf{R}^{p \times q}[\sigma]$.

Proposition A.2: A set $\mathcal{B} \in l_2^q$ is an l_2 -system if and only if $\mathcal{B} = \{w \in l_2^q; R(\sigma)w = 0\}$ for some $R \in \mathbf{R}^{p \times q}[s]$ and $p \in \mathbf{N}$.

Proof: The if-part is trivial, and the other part is proved by construction; see [10, Theorem 5]. \square

The complexity of a system is measured in terms of its dimension. Considered as a linear space, every nonzero l_2 -system has infinite dimension. Therefore we consider its dimension on finite time intervals.

Definition A.3: For a given system \mathcal{B} let $\mathcal{B}^0 := \{w \in \mathcal{B}; w(t) = 0 \text{ for } t < 0\}$, and let \mathcal{B}_0^0 be the restriction of \mathcal{B}^0 to time $t = 0$.

- 1) The rank of \mathcal{B} is defined as $m(\mathcal{B}) := \dim(\mathcal{B}_0^0)$.
- 2) The degree of \mathcal{B} is defined as $n(\mathcal{B}) := \dim(\mathcal{B}|_{[0, \infty)}) \bmod \mathcal{B}_0^0|_{[0, \infty)})$.

The rank and degree determine the dimension of an l_2 -systems on finite intervals. To be specific, if \mathcal{B} is a system with rank m and degree n , then $\dim(\mathcal{B}|_T) = mN + n$ for intervals $T \subset \mathbf{Z}$ of length $N \geq n$. The rank and degree have the following interpretation. The rank is the number of degrees of freedom for a system at each time instant, given the past. This is equal to the number of inputs in the system. The degree measures the remaining freedom due to initial conditions. This is equal to the number of states. This is made precise in the following result.

Proposition A.4: The rank and degree of a system equal, respectively, the number of auxiliary inputs and the number of states in a minimal state representation.

Proof: For a proof we refer to [10, Theorem 9]. \square

APPENDIX B

PROOFS

Proof of Proposition 4.3:

Necessity of 1: Let $\mathcal{R} := \text{im}[B \ AB \ \dots \ A^{n-1}B]$, then $A\mathcal{R} \subset \mathcal{R}$ and $\text{im } B \subset \mathcal{R}$. If (A, B) is not controllable there is a choice of basis for the state space such that with respect to this basis A and B take the form $\begin{bmatrix} A_1 & A_3 \\ 0 & A_2 \end{bmatrix}$ and $\begin{bmatrix} B_1 \\ 0 \end{bmatrix}$.

So in a corresponding partition $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ of the state variables it holds that $\sigma x_2 = A_2 x_2$. As this admits only the zero solution in l_2 , x_2 can be removed.

Necessity of 2: Suppose there exists an F such that $(A + BF, C + DF)$ is not observable. Then the unobservable components of the state can be removed, as in the previous part.

Necessity of 3: An obvious condition is that $\begin{bmatrix} B \\ D \end{bmatrix}$ has full column rank, so it suffices to prove that $\ker D \subset \ker B$. Suppose $\ker D \setminus \ker B \neq \{0\}$, then there exists an invertible $R \in \mathbf{R}^{m \times m}$ such that $\begin{bmatrix} BR \\ DR \end{bmatrix} = \begin{bmatrix} B' & b \\ D' & 0 \end{bmatrix}$ with $0 \neq b \in \mathbf{R}^n$. Then \mathcal{B} is represented by $\sigma x = Ax + B'v' + bz$ and $w = Cx + D'v'$. As z influences w only with a delay, we remove its direct influence on σx by defining $\sigma x' := \sigma x - bz$. This gives $\sigma x' = Ax' + B'v' + Ab\sigma^{-1}z$ and $w = Cx' + D'v' + Cb\sigma^{-1}z$. Hence (A, B, C, D) is equivalent to $(A, [B' \ Ab], C, [D' \ Cb])$. From Proposition 4.2 it follows that this is equivalent to $(A + Abf, [B' \ Ab], C + Cbf, [D' \ Cb])$, for all $f \in \mathbf{R}^{1 \times n}$. As $b \neq 0$, f can be chosen such that $I_n + bf$ is singular, by taking $f = -b^T/\|b\|^2$. It is easily verified that then $(A + Abf, C + Cbf) = (A(I_n + bf), C(I_n + bf))$ is not observable. From the Necessity of 2 it follows that the state dimension can be reduced.

Sufficiency of the Conditions: Let (A, B, C, D) be an SR with m auxiliary inputs and n states that satisfies Conditions 1), 2), and 3). From controllability and Proposition 4.2 it follows that without loss of generality we may assume that A is asymptotically stable. We prove that the rank of \mathcal{B} equals m , and its degree equals n . Then the result follows from Proposition A.4.

Concerning the rank, consider \mathcal{B}^0 as introduced in Definition A.3. Observability implies that trajectories $w \in \mathcal{B}^0$ have state zero at time $t = 0$. So $\mathcal{B}_0^0 = \text{im } D$, and from Condition 3) it follows that $\dim \mathcal{B}_0^0 = \text{rank } D = m$. Concerning the degree, we consider the space $\mathcal{B}|_{[0, \infty)} \bmod \mathcal{B}_0^0|_{[0, \infty)}$; cf. Definition A.3-2). This space can be parameterized by the initial state x_0 . Observability implies that this parameterization is injective, from which the result follows. \square

Proof of Proposition 5.2: Write $(A', B', C', D') := (S(A + BF)S^{-1}, SBR, (C + DF)S^{-1}, DR)$. Then

$$\begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix} = \begin{pmatrix} S & 0 \\ 0 & I_m \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} S^{-1} & 0 \\ FS^{-1} & R \end{pmatrix}.$$

Equation (13) for (A', B', C', D') gives, with $K = S^T S$

$$\begin{aligned} & \begin{pmatrix} A^T & C^T \\ B^T & D^T \end{pmatrix} \begin{pmatrix} K & 0 \\ 0 & I_m \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} \\ &= \begin{pmatrix} S^{-1} & 0 \\ FS^{-1} & R \end{pmatrix}^{-T} \begin{pmatrix} S^{-1} & 0 \\ FS^{-1} & R \end{pmatrix}^{-1} \\ &= \begin{pmatrix} S^T & -F^T R^{-T} \\ 0 & R^{-T} \end{pmatrix} \begin{pmatrix} S & 0 \\ -R^{-1}F & R^{-1} \end{pmatrix} \\ &= \begin{pmatrix} K + F^T(RR^T)^{-1}F & -F^T(RR^T)^{-1} \\ -(RR^T)^{-1}F & (RR^T)^{-1} \end{pmatrix}. \end{aligned}$$

Now verification of (16), (17), and (14) is straightforward. For a proof of the uniqueness of K we refer to [8]. \square

Proof of Proposition 5.3: From (13) it follows that $A^T A = I_n - C^T C$, hence $|Ax|^2 = |x|^2 - |Cx|^2 \leq |x|^2$. So A is stable. We prove that the representation is not minimal if A is not asymptotically stable. In that case A has an eigenvalue λ with $|\lambda| = 1$. Let x denote a corresponding eigenvector, and x^* its complex conjugate. Then $|Cx|^2 = x^* C^T C x = x^* x - x^* A^T A x = |x|^2 - |Ax|^2 = 0$. This implies that $CA^k x = 0$, $k \geq 0$, so that (A, C) is not observable and hence not minimal.

From the proof of Proposition 5.2 it follows that (15)–(17) are necessary conditions. Equation (15) determines S modulo a left unitary factor, corresponding to U . Equation (16) determines R modulo a right unitary factor, corresponding to V . \square

Proof of Theorem 6.2: From Definition 5.1 it follows that $(A, \tilde{B}, C, \tilde{D})$ as defined in part 4) is an ISR. As A is asymptotically stable [see Proposition 5.3-1)] the corresponding image representation $\tilde{G}: l_2^{q-m} \rightarrow l_2^q$ is well defined. We first prove that

$$GG^* + \tilde{G}\tilde{G}^* = I_q. \quad (28)$$

Define $\bar{B} = [B \ \tilde{B}]$ and $\bar{D} = [D \ \tilde{D}]$. Then (A, \bar{B}, C, \bar{D}) is also an isometric state representation. Let $\bar{G}(\sigma): l_2^q \rightarrow l_2^q$ denote the corresponding image representation. Then $\bar{G} = [G \ \tilde{G}]$, so $GG^* + \tilde{G}\tilde{G}^* = \bar{G}\bar{G}^*$. It remains to prove that $\bar{G}\bar{G}^* = I_q$, or equivalently, that $\bar{G}^* = \mathcal{R}_q \bar{G}^T \mathcal{R}_q$ is isometric. The operator \mathcal{R}_q is clearly isometric, and because \bar{G}^T is the image representation induced by the ISR $(A^T, C^T, \bar{B}^T, \bar{D}^T)$, it is also isometric. This proves (28).

- 1) From Theorem 6.1 it follows that $\tilde{G}\tilde{G}^* w$ is the optimal approximation of w in $\mathcal{B}(A, \tilde{B}, C, \tilde{D})$. From (28) we obtain $\{w \in \mathcal{B}(A, \tilde{B}, C, \tilde{D})\} \Leftrightarrow \{\tilde{G}\tilde{G}^* w = w\} \Leftrightarrow \{GG^* w = 0\} \Leftrightarrow \{w \in \mathcal{B}^\perp\}$. So $\text{im } \tilde{G} = \mathcal{B}^\perp$, which proves 1).
- 2) Equation (28) shows that every $w \in l_2^q$ can be decomposed into a part contained in \mathcal{B} and one in \mathcal{B}^\perp .
- 3) As $(A, \tilde{B}, C, \tilde{D})$ is a state representation of \mathcal{B}^\perp , it follows that $m(\mathcal{B}^\perp) \leq q-m$ and $n(\mathcal{B}^\perp) \leq n$. From part 2) it follows that $m(\mathcal{B}^\perp) \geq q-m$, so $m(\mathcal{B}^\perp) = q-m$. As $(\mathcal{B}^\perp)^\perp = \mathcal{B}$, there holds that $n = n((\mathcal{B}^\perp)^\perp) \leq n(\mathcal{B}^\perp)$, so that $n(\mathcal{B}^\perp) = n$. This proves 3).
- 4) That $(A, \tilde{B}, C, \tilde{D})$ is an ISR of \mathcal{B}^\perp was proved in 1), and minimality follows from 3) and Proposition A.4.
- 5) Equation (28) implies that $\tilde{w} = w - \hat{w} = w - GG^* w = \tilde{G}\tilde{G}^* w$, and the result follows from Theorem 6.1. \square

Proof of Theorem 8.1: It suffices to prove the theorem for a minimal isometric SR. This can be seen as follows. If condition 2) holds for an arbitrary minimal SR, then it holds for all equivalent minimal SR's, as in the transformation $(S(A + BF)S^{-1}, SBR, (C + DF)S^{-1}, DR)$, the auxiliary input and state are linearly transformed to $R^{-1}(\hat{v} - F\hat{x})$ and $S\hat{x}$; see the proof of Proposition 4.2.

So let (A, B, C, D) be a minimal ISR of \mathcal{B} , and let \tilde{B}, \tilde{D} , be defined as in Theorem 6.2-4). In the proof we will make use of the following relations, which follow from the projection

scheme in Section VI; see Fig. 2.

$$\text{for } \hat{v}: x = A^T \sigma x + C^T w, \quad \hat{v} = B^T \sigma x + D^T w \quad (29)$$

$$\text{for } \hat{v}: x = A^T \sigma x + C^T w, \quad \hat{v} = \tilde{B}^T \sigma x + \tilde{D}^T w \quad (30)$$

$$\text{for } \hat{w}: \sigma \hat{x} = A \hat{x} + B \hat{v}, \quad \hat{w} = C \hat{x} + D \hat{v} \quad (31)$$

$$\text{for } \hat{w}: \sigma \hat{x} = A \hat{x} + \tilde{B} \hat{v}, \quad \hat{w} = C \hat{x} + \tilde{D} \hat{v}. \quad (32)$$

Further, from (13) we obtain

$$\begin{aligned} \hat{x} &= A^T \sigma \hat{x} + C^T \hat{w}, & \hat{v} &= B^T \sigma \hat{x} + D^T \hat{w}, \\ 0 &= \tilde{B}^T \sigma \hat{x} + \tilde{D}^T \hat{w} \end{aligned} \quad (33)$$

$$\begin{aligned} \hat{x} &= A^T \sigma \hat{x} + C^T \hat{w}, & 0 &= B^T \sigma \hat{x} + D^T \hat{w}, \\ \hat{v} &= \tilde{B}^T \sigma \hat{x} + \tilde{D}^T \hat{w}. \end{aligned} \quad (34)$$

We first describe those optimality conditions that can be derived straightforwardly from the model improvement constructions in Theorem 7.2.

Lemma B.1:

- 1) Construction 1 gives no improvement if and only if $\text{cov}([\hat{x}, \hat{v}], \hat{w}) = 0$.
- 2) Construction 2 gives no improvement if and only if $\text{cov}(\hat{v}, [\sigma \hat{x}, \hat{w}]) = 0$.
- 3) If Construction 3 gives no improvement, then $\text{cov}(\hat{v}, \hat{v}) = 0$.

Proof of the Lemma: We use the notation of Theorem 7.2.

- 1) As $\hat{w} = C\hat{x} + D\hat{v}$, clearly $\hat{w} \in \mathcal{E}$. Construction 1 does not yield an improvement iff the projection of \hat{w} onto \mathcal{E} is zero, from which the result follows.
- 2) Construction 2 gives no improvement iff the projection of \hat{w} onto F is zero, cf. part 1). This is equivalent to $\langle \hat{w}, (C(\sigma I - A)^{-1} B' + D') \hat{v} \rangle = 0$ for all B', D' . From (34) this is equivalent to $\langle \hat{w}, \sum_{k=1}^{\infty} C A^{k-1} B' \sigma^{-k} \hat{v} + D' \hat{v} \rangle = \langle \sum_{k=1}^{\infty} B'^T (A^T)^{k-1} C^T \sigma^k \hat{w} + D'^T \hat{w}, \hat{v} \rangle = \langle B'^T \sigma \hat{x} + D'^T \hat{w}, \hat{v} \rangle = 0$, from which the result follows.
- 3) If $\text{cov}(\hat{v}, \hat{v}) \neq 0$, then construction 3 decreases $\|\hat{v}\| = \|\hat{w}\|$. \square

Returning to Theorem 8.1, we first prove part 2. From the lemma it follows that $\text{cov}(\hat{v}, \hat{v}) = 0$, and $\text{cov}(\hat{v}, \hat{x}) = \text{cov}(\hat{v}, A^T \sigma \hat{x} + C^T \hat{w}) = 0$. As \tilde{D} is injective, $\text{cov}(\hat{x}, \hat{v}) = 0$ if and only if $\text{cov}(\hat{x}, \tilde{D} \hat{v}) = 0$, which is equivalent to $\text{cov}(\hat{x}, \hat{w} - C \hat{x}) = -\text{cov}(\hat{x}, C \hat{x}) = 0$. So to prove 2 it remains to show that $\text{cov}(\hat{x}, \hat{x}) = 0$. As (A^T, B^T) is observable, this is equivalent to $\text{cov}(\sigma \hat{x}, B^T A^{T^k} \sigma \hat{x}) = 0$ for all $k \geq 0$. We prove this by induction. For $k = 0$, the lemma shows that $\text{cov}(\sigma \hat{x}, B^T \sigma \hat{x}) = \text{cov}(\sigma \hat{x}, -D^T \hat{w}) = 0$. Now suppose that $\text{cov}(\sigma \hat{x}, B^T A^{T^k} \sigma \hat{x}) = 0$ for $k \leq N$. Then $\text{cov}(\sigma \hat{x}, B^T A^{T^{N+1}} \sigma \hat{x}) = \text{cov}(A \hat{x} + B \hat{v}, B^T A^{T^{N+1}} \sigma \hat{x}) = \text{cov}(A \hat{x}, B^T A^{T^{N+1}} \sigma \hat{x}) = \text{cov}(A \hat{x}, B^T A^{T^N} (\hat{x} - C^T \hat{w})) = A \text{cov}(\hat{x}, B^T A^{T^N} \hat{x}) = 0$. This proves Theorem 8.1.2. Concerning the equivalence of 1, 2, and 3, the implications $3 \Rightarrow 2 \Rightarrow 1$ are trivial. Further, if 1 holds then $\text{cov}(\hat{x}, \hat{x}) = \text{cov}(\sigma \hat{x}, \sigma \hat{x}) = \text{cov}(A \hat{x} + B \hat{v}, A \hat{x} + \tilde{B} \hat{v}) = \text{cov}(A \hat{x}, A \hat{x}) = A \text{cov}(\hat{x}, \hat{x}) A^T$. As A is asymptotically stable, it follows that $\text{cov}(\hat{x}, \hat{x}) = 0$, so 1 implies 2. Finally, 3 is easily derived from 2 by using (31) and (32). \square

Proof of Theorem 8.2: Let w be a given observation and let \hat{w} be its optimal approximation in \mathcal{B} . Further let

(A, B, C, D) be a minimal SR of \mathcal{B} with A asymptotically stable, and write the approximation error as $\tilde{w} := w - \hat{w} = w - (C(\sigma I - A)^{-1}B + D)\hat{v}$. We have to prove that \mathcal{B} satisfies the optimality conditions if and only if the derivative of $\|\tilde{w}\|$ with respect to the parameters in A, B, C, D , and \hat{v} is zero. First we analyze the tangent space of \tilde{w} with respect to these parameters. Let \mathcal{E} and \mathcal{F} be defined as in Theorem 7.2, and let $\mathcal{G} := \{\bar{w} \in l_2^q; \exists H \in \mathbb{R}^{n \times n} \text{ such that } \bar{w} = C(\sigma I - A)^{-1}H\hat{x}\}$ with \hat{x} the state corresponding to \hat{w} .

Lemma B.2: The tangent space \mathcal{T} of $\tilde{w} = w - (C(\sigma I - A)^{-1}B + D)\hat{v}$ is given by $\mathcal{T} = \mathcal{B} + \mathcal{E} + \mathcal{F} + \mathcal{G}$.

Proof of the Lemma: The tangent space is defined as the smallest closed subspace of l_2 containing all partial derivatives. Note that \tilde{w} is linear in \hat{v} , in B , in C and in D . A change of \hat{v} corresponds to adding $\bar{w} \in \mathcal{B}$ to \tilde{w} , a change of C and D to adding $\bar{w} \in \mathcal{E}$ to \tilde{w} and a change of B and D to adding $\bar{w} \in \mathcal{F}$ to \tilde{w} . It remains to prove that the derivatives of \tilde{w} with respect to the parameters in A span the space \mathcal{G} . For $H \in \mathbb{R}^{n \times n}$ let x' be defined by $\sigma x' = (A - H)x' + B\hat{v}$ and let $w' := Cx' + D\hat{v}$. The corresponding error is $\tilde{w}' := w - w'$, so that the change in \tilde{w} is given by $\bar{w} = \tilde{w}' - \tilde{w} = \hat{w} - w' = C(\hat{x} - x') = C\bar{x}$ for $\bar{x} := \hat{x} - x'$. As $\sigma\bar{x} = A\bar{x} - (A - H)x' = A\bar{x} + Hx' = A\bar{x} + H\hat{x} - H\bar{x}$, ignoring the second-order term $H\bar{x}$ for small H gives the result. \square

We will next prove the theorem by showing that the optimality conditions and the stationarity condition are both equivalent to $\tilde{w} \perp \mathcal{T}$.

Stationarity is equivalent to the condition that $\lim_{\delta \rightarrow 0} \delta^{-1} \{\|\tilde{w} + \delta\bar{w}\| - \|\tilde{w}\|\} = 0$ for all $\bar{w} \in \mathcal{T}$. It is easily verified that this limit equals $\langle \tilde{w}, \bar{w} \rangle / \|\bar{w}\|$, so stationarity is equivalent to $\tilde{w} \perp \mathcal{T}$.

Finally we show that $\tilde{w} \perp \mathcal{T}$ is equivalent to the optimality conditions. First, suppose that the optimality conditions hold. As \hat{w} is an optimal approximation within \mathcal{B} there holds that $\tilde{w} \perp \mathcal{B}$. Further, Theorem 8.1-3) states that $\text{cov}([\hat{v}, \hat{x}], \tilde{w}) = 0$, so that $\tilde{w} \perp \mathcal{E}$, and $\text{cov}(\hat{v}, [\sigma\hat{x}, \tilde{w}]) = 0$, so that the proof of Lemma B.1-2) shows that $\tilde{w} \perp \mathcal{F}$. Finally, for $\bar{w} \in \mathcal{G}$ given by $\bar{w} = \sum_{k=1}^{\infty} CA^{k-1}H\sigma^{-k}\hat{x}$ we obtain by using (34) that $\langle \bar{w}, \tilde{w} \rangle = \langle \sum_{k=1}^{\infty} CA^{k-1}H\sigma^{-k}\hat{x}, \tilde{w} \rangle = \langle \hat{x}, \sum_{k=1}^{\infty} H^T A^{k-1} C^T \sigma^k \tilde{w} \rangle = \langle \hat{x}, H^T \sigma \tilde{x} \rangle = 0$, so $\tilde{w} \perp \mathcal{G}$. From Lemma B.2 it follows that $\tilde{w} \perp \mathcal{T}$.

Second, supposing that $\tilde{w} \perp \mathcal{T}$ we prove the optimality conditions. The fact that $\tilde{w} \perp \mathcal{E} + \mathcal{F}$ implies the conditions in Lemma B.1-2) and B.1-2), cf. the proof of that lemma. The condition in Lemma B.1-3) follows from that in B.1-2) by using (34). Further, the optimality conditions were derived from these conditions in the proof of Theorem 8.1. \square

Proof of Proposition 8.4: First we prove that \mathcal{Z} is a linear space. Observe that for every $z \in \mathcal{B}^\perp$ the optimal approximation of $\hat{w} + z$ within \mathcal{B} is given by \hat{w} . Let (\hat{v}, \hat{x}) be the auxiliary input and state for \hat{w} in a minimal SR of \mathcal{B} , and let $(\tilde{v}_z, \tilde{x}_z)$ be defined analogously for z in \mathcal{B}^\perp . As (\hat{v}, \hat{x}) is fixed, the condition in Theorem 8.1-2) for stationarity of \mathcal{B} with respect to $\hat{w} + z$ consists of linear restrictions on $(\tilde{v}_z, \tilde{x}_z)$. As \tilde{x}_z is a linear function of \tilde{v}_z , these conditions can be expressed as linear restrictions on \tilde{v}_z alone. This shows that \mathcal{B} is a stationary point for $\hat{w} + z$ if and only if the

auxiliary input \tilde{v}_z is restricted to a linear subspace of l_2^{q-m} , from which the linearity of \mathcal{Z} follows. Next we show that the minimum in (22) is achieved by taking $\bar{w}_0 = \tilde{w} - \tilde{w}'$. As $w - \bar{w}_0 = \hat{w} + \tilde{w} - (\tilde{w} - \tilde{w}') = \hat{w} + \tilde{w}'$ and $\tilde{w}' \in \mathcal{Z}$, it follows that \mathcal{B} is stationary for $w - \bar{w}_0$, by definition of \mathcal{Z} . Further, $\bar{w} \in \mathcal{B}^\perp$ is such that \mathcal{B} is stationary for $w - \bar{w} = \hat{w} + \tilde{w} - \bar{w}$ if and only if $z := \tilde{w} - \bar{w} \in \mathcal{Z}$. Now the norm of $\bar{w} = \tilde{w} - z$ is minimized by taking $z = \tilde{w}'$, the orthogonal projection of \tilde{w} on \mathcal{Z} , hence $\bar{w}_0 = \tilde{w} - \tilde{w}'$. \square

REFERENCES

- [1] P. E. Caines, *Linear Stochastic Systems*. New York: Wiley, 1988.
- [2] K. Glover, "All optimal Hankel norm approximations of linear multivariable systems and their L^∞ error bounds," *Int. J. Contr.*, vol. 39, no. 6, pp. 1115-1193, 1984.
- [3] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1983.
- [4] E. J. Hannan and M. Deistler, *The Statistical Theory of Linear Systems*. New York: Wiley, 1988.
- [5] C. Heij, *Deterministic Identification of Dynamical Systems* (Lecture Notes in Control and Information Sciences), vol. 127. Berlin: Springer-Verlag, 1989.
- [6] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [7] B. C. Moore, "Principal component analysis in linear systems: Controllability, observability, and model reduction," *IEEE Trans. Automat. Contr.*, vol. AC-26, no. 1, pp. 17-32, 1981.
- [8] H. J. Payne and L. M. Silverman, "On the discrete time algebraic Riccati equation," *IEEE Trans. Automat. Contr.*, vol. AC-18, no. 3, pp. 226-234, 1973.
- [9] J. T. Spanos, M. H. Milman, and D. L. Mingori, "A new algorithm for L_2 optimal model reduction," *Automatica*, vol. 28, no. 5, pp. 897-909, 1992.
- [10] J. C. Willems, "From time series to linear system, part I: Finite dimensional linear time invariant systems; Part II: Exact modelling; part III: Approximate modelling," *Automatica*, vol. 22/23, nos. 5, 6, 1, pp. 561-580, 675-694, 87-115, 1986, 1987.
- [11] —, "Paradigms and puzzles in the theory of dynamical systems," *IEEE Trans. Automat. Contr.*, vol. 36, no. 3, pp. 259-294, 1991.
- [12] J. C. Willems and C. Heij, "Scattering theory and approximation of linear systems," *Modeling, Identification and Robust Control*, C. Byrnes and A. Lindquist, Eds. Amsterdam: North-Holland, 1986, pp. 397-411.



Christiaan Heij (M'88) was born in Arnhem, The Netherlands, in 1957. He studied econometrics, mathematics, and philosophy and received the Ph.D. degree from the University of Groningen, in 1988, for a dissertation on identification of linear systems.

Since 1989 Dr. Heij has been an Assistant Professor at the Econometric Institute of the Erasmus University in Rotterdam. His research interests are in the area of modeling and identification, in particular the theory and applications of linear systems and econometrics.



Berend Roorda was born in Assen, The Netherlands, in 1965. He obtained the M.Sc. degree from the Department of Mathematics of the University of Groningen, The Netherlands, in 1989. He is currently finishing his Ph.D. dissertation on approximate modeling by the global total least squares method.

He was researcher at the PTT Research Telematics Laboratory in Groningen in 1989. Since 1990 he has been working as a Research Assistant at the Tinbergen Institute and the Econometric Institute of

the Erasmus University Rotterdam.